

# Reading hybrid texts: Remarks on text/image transitions

A Baptista<sup>1</sup>, I Faria<sup>2</sup> & P Luegi<sup>2</sup>

<sup>1</sup>ESE - Instituto Politécnico do Porto

<sup>2</sup>Laboratório de Psicolinguística – FLUL

<sup>1</sup>acbaut@vizzavi.pt

## ABSTRACT

The reading of printed materials implies the visual processing of information originated in two distinct semiotic systems. The rapid identification of redundancy, complementation or contradiction rhetoric strategies between the two information types may be crucial for an adequate interpretation of bimodal materials. Hybrid texts (verbal and visual) are particular instances of bimodal materials, where the redundant information is often neglected while the complementary and the contradictory ones are essential. Studies using the 504 ASL eye-tracking system while reading either additive or exhibiting captions (Baptista, 2009) revealed fixations on the verbal material and transitions between the written and the pictorial in a much higher number and duration than the initially foreseen as necessary to read the verbal text. We therefore hypothesized that confirmation strategies of the written information are taking place, by using information available in the other semiotic system. Such eye-gaze patterns obtained from denotative texts and pictures seem to contradict some of the scarce existing data on visual processing of texts and images, namely cartoons (Carroll, Young and Guertain, 1992), descriptive captions (Hegarty, 1992 a and b), and advertising images with descriptive and explanatory texts (cf. Rayner and Rotello, 2001, who refer to a previous reading of the whole text before looking at the image, or even Rayner, Miller and Rotello, 2008 who refer to an earlier and longer look at the picture) and seem to consolidate findings of Radach et al. (2003) on systematic transitions between text and image. By framing interest areas in the printed pictorial material of non redundant hybrid texts, we have identified the specific areas where transitions take place after fixations in the verbal text. The way those transitions are processed brings a new interest to further research.

## 1. INTRODUCTION

Nowadays the reading of printed materials (as visual/verbal rhetorical elaborated texts) often implies the visual processing of information originated in two distinct semiotic systems. We call those texts hybrid texts<sup>1</sup> when the semantic output of the annexation of those two kinds of texts is not a simple addition of their semantic properties, but a factorial product (as it happens, for instance, in some examples of illustrated texts, some advertisements and works of art). In order to understand hybrid texts the reader must cognitively process the semantic difference or similitude between the mental images the text allows him to build and the image he is really seeing and he must realise the significance of the annexation of that particular image to that particular text.

In hybrid texts we may have different situations. Sometimes images do increase immediately the exhibiting effect of the texts (the capability that verbal rhetorical strategies in descriptive, narrative,

---

<sup>1</sup> This concept of hybrid text, proposed in Baptista, 2009, within the visual processing analysis of captioned images presupposes a dynamic concept of text (cf. Scherner, 2000), in which the meaning of the global text is supported by each of the textual instances and the visual perception itself is guided based on empirical and culturally acquired knowledge. In this dynamic conception, the expectations generated by the reader will be progressively validated or invalidated, thus determining the way in which reading and understanding should proceed; with hybrid texts, the process of constructing expectations about meanings heavily depends on the way in which processing is also achieved in a hybrid manner.

explicative or expositive texts can have to develop mental images in reader's mind), sometimes they do not. When images are rhetorically complex (mainly when they show some kind of alotopic information or when pictorial elements neither predictably match with the background knowledge of the reader or their expectations, nor can they be understood in a denotative way) they may convey a lot of conflicting meanings that might reduce the centrality effect of attention on the verbal text or otherwise strongly direct the attention to text in an attempt to clarify the image with the verbal information supplied. Sometimes the text seems not sufficient to clarify the image. In those cases the cognitive processing of the whole meaning of the text may depend on the visual processing strategies. While trying to understand the whole meaning of the hybrid text the reader can exhibit a particular visual scanpath.

Understanding an image is not limited to detecting isomorphic and disomorphic characteristics of objects or events it represents or to building a mental representation for its information content, but it also implies the identification of the objectives of its reproduction and the detection of semiotic characteristics within its rhetorical organization.

## 2. READING HYBRID TEXTS

We will assume that understanding the image of a hybrid text presupposes more than a level of analysis and that the identification of a representative dimension, despite implying the identification/understanding of certain structural characteristics, is not enough. One must thus look for the main characteristics of visual processing for hybrid texts. As the drowning of attention in first place or the duration of fixations during the visual processing of the hybrid message cannot be taken as the right clues to identify how the cognitive processing is being done, we also propose to consider the type and number of transitions between texts and image.

The amount and the type of information those two systems present can be similar, complementary or opposite. In hybrid texts the main strategies are neither redundancy nor simple illustration (these strategies being more common in mixed texts). Hybrid texts (verbal and visual) are particular instances of bimodal materials, where the predictable and redundant information is often neglected while the complementary and the contradictory information is essential for the comprehension of the message.

Therefore, the rapid identification of redundancy, complementarity, expansion or contradiction rhetoric strategies between the two information types may be crucial for an adequate interpretation of hybrid materials and for the reduction of possible situations of communicative entropy.

Despite the eventual competition on grasping our eyes' attention, we propose that text and image establish a partnership relation and not a real opposition or resistance (cf. Mitchell, 1994); in fact, we would like to call it a rhetorical attraction relation. When image and text share this kind of rhetoric relation, gathering for a sequential graphic order (that may link syntactically words and images) is not important and we want to focus our attention on the search of a semantic relation that our eyes can help to perform.

As we proposed for captions (cf. Baptista, 2009), this attraction between text and image might be performed by a focalization operation led by one of the textual instances over the other. This focalization will conduct the visual observation and might create a perceptual and semantic scanpath showing how some information in the text can be underlined or neglected by the action of either the text or the image.

## 3. VISUAL PROCESSING OF IMAGES AND TEXTS

There is a strong relationship between the eyes and the mind, and the processing of continuous information while the ocular movements are programmed. As part of understanding or not understanding the visually processed data, the direction of the movements can be altered and the initial fixations can determine the direction of those that follow<sup>2</sup> (Becker and Jurgens, 1979).

---

<sup>2</sup> This data explains the existence of regressive movements during the reading of verbal texts. These can take place when something in the text is not understood by the reader, thus inverting the progressive sequential order of eye movements.

As a result it is important to determine what successively attracts our gaze during the reading of an image, given that each place of fixation can be determined by the preceding fixation. Some relevant data concerning the visual processing of images and texts in a single perception area can be recalled:

- the eyes quickly move towards more informative regions of an image (Antes, 1974 and Mackworth and Morandi, 1979);
- during the process of observing an image, its fulcral element is identified at the beginning of observation and very quickly, on average during the first two fixations, while remaining fixations complete the details of the image (Loftus and MacKworth, 1978);
- in a given scenario, the eyes quickly move towards an object when outside of its predicted location (Friedman, 1979; Friedman and Liebelt, 1981) which may explain the speed of alotopic understanding.
- when the eye is stuck on an image, it tends to move vertically to the text information placed right beneath (Brandt, 1945: 90-93);
- while reading cartoons with captions there are no repetitive movements between text and image, the image is normally read after the text, so the processing of image and text are relatively isolated actions (Carroll, Young and Guertain, 1992)
- in text and image interactions the cognitive reading and understanding process is highly directed to the text (Hegarty, 1992 a e b).
- on bimedia advertising, subjects first observe texts, spending more time on them, and only then the images (Rayner, Rotello et al., 2001). In this Rayner and Rotello's study (2001) even though the authors consider the integration of information contained in the text as fundamental to understanding the images tested, we do not consider this a factual approach. They have not defined the characteristics of images and texts and did not distinguish the functions of interaction between the two, since one of the images tested (from car publicity) seems to us to be merely illustrative, and the other, interpretive representative. Rayner, Miller and Rotello (2008) find otherwise that initial and longer fixations can be registered on images, depending on the goal of the observation instruction;
- important or interesting objects (in the majority of these works, the significant element is the *out-of-place* element which we will refer to as alotopic element) are the target of a greater number and longer fixations than isotopic objects of little interest (Rayner and Pollatsek's, 1989); Underwood and Foulsham (2006) alert to the consistency of the incongruent effect besides the visual salience of an object in the analysis of earlier fixations;

### 3.1. Transitions between written and pictorial instances in captioned images

As a sub experiment in a large work about the importance of different types of captions in the cognitive processing of information conveyed by an image and its caption, we have conducted an experiment (cf. Baptista, 2009) with 9 adult subjects (using the Eyenal software from the ASL 504 eye-tracking system monitoring the observation of captioned images) from which we got information about the visual processing in the hybrid image/caption texts patterns. The scanpath obtained presented systematic crossed fixations between images and captions of different kinds (additive, exhibiting, additive +exhibiting) referred as B1 e B2.

In spite of pre-referred data from several authors (cf. Carroll, Young and Guertain (1992); Hegarty1 (1992 a e b) Rayner, Rotello et al. (2001); we verified (with the exception of three cases) that all the subjects reading both the image and the caption made transitions between both areas while reading (two subjects have not read the caption at all). The scanpath shows that patterns of visual processing register various crossing points between image and caption, when distributed by two distinct framed areas; the majority of participants neither observe the image in its totality, nor read the caption in a row, distributing attention in an interpolated manner over the two areas of interest, with the majority beginning with the image; the same subject observing different images makes a different number of transitions (Table 1).

**Table 1.** *Number and sequence of transitions between images and captions.*

Distribution of transitions																		
image	Subjects	off	image	caption	image	caption	image	caption	image	caption	image	caption	image	caption	image	caption	image	off
Lamp exhibiting caption	6L	1		2	3													4
	7L	1		2	3	4	5	6										7
	8L		1	2	3	4	5											6
	9L	1	2	3	4													5
Stamp exhibiting caption	6L	1	2	3	4	5		7		9	11	12						13
	7L	1		2	3													4
	8L	1	2	3	4	5	6	7										8
	9L	1	2	3	4	5	6	7	8									9
Writing Case exhibiting caption	6L			1		3	4											5
	7L	1	2	3	4	5	6	7										8
	8L	1	2	3	4	5	6	7										8
	9L	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Ring Box additive + exhibiting caption	6L	1		2	3		5	6		8		10						11
	7L	1		2	3	4	5	6	7	8								9
	8L			1	2	3	4	5	6									7
	9L	1	2															
Pipe additive + exhibiting caption	1L	1	2	3														4
	2L	1	2	3	4													5
	3L		1	2	3													4
	4L	1	2	3	4													5
Keys additive + exhibiting caption	5L	1	2	3	4	5												6
	6L	1	2	3		5	7	8										9
	7L		1	2	3	4	5	6	7									8
	8L	1	2	3	4	5												6
Stones additive + exhibiting caption	9L	1	2	3	4													5
	1L	1	2	3	4													5
	2L	1	2	3	4													5
	3L	1	2															3
	4L	1	2	3	4													5
	5L	1	2	3	4	5	6											7

In a qualitative analysis, through a quite informal but significant scanpath observation we found that several subjects, in several images, made transitions from the caption to the image area where the objects (or parts of them) referred to on the caption are located, apparently accomplishing a confirmation path for the verbal information read. We also verified that subjects do not approach different captioned images always in the same way; sometimes they begin reading the caption, other times the image (Table 2).

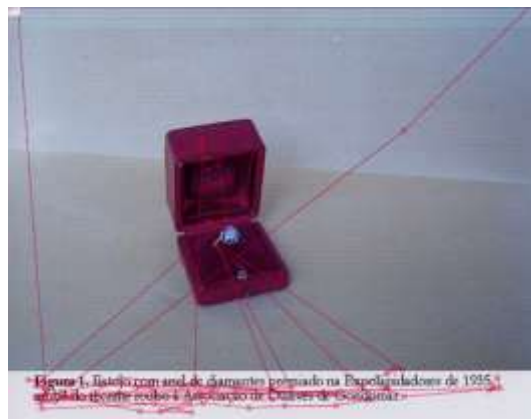
**Table 2.** *Fixations and transitions between images and captions*

Subjects	First fixation on text	First fixation on image	At least 3 transitions	More than 3 transitions
28	5	23	5	16

Figure 1 and Figure 2 are two examples showing the scanpaths of two different subjects reading movements over the captioned picture, revealing the transitions between an image and its block caption

where we can see that the observers do not observe completely either the whole caption, or the whole image at once, but they distribute their attention between those two areas of interest.

That transition data gathered from captioned images with quite denotative images and some redundant information degree between text and image developed the need to see how transitions could look like if there was no redundant information between both instances of the hybrid text and if the rhetoric complexity of the image was significant.



**Figure 1.** Transitions between caption and text. Ring box



**Figure 2.** Transitions between caption and text. Writing Chinese case.

### 3.2. Transitions between written and pictorial instances in a hybrid advertising text

To study transitions text/image we collected more information from another experiment with a hybrid advertising text where the main objective was to observe the processing of a rhetorically elaborated hybrid advertising text formed by a complex connotative image and a very short text, the brand designation, considering the information recall and its comprehension. In addition we conducted an experiment to observe the visual processing of the whole hybrid text from which we collected transitions data.

#### 3.2.1 Experiment.

We have used an advertising image which, despite being rhetorically complex and elaborated according to unequivocally connotative parameters, is potentially interpreted without text and where the basically contextualizing text, while significantly contributing to understanding the connotative senses of the pictorial text, does not describe nor affect the increase in image meanings.

#### 3.2.2 Materials.

We chose an image with little text (just the brand designation written in white lower case letters) located in the lower right hand corner of the image (an area which attracts little significant attention), accompanied by an icon in white, the head of a snake surrounded by a red field. It is an allegorical visual text rich in metonyms, metaphors, synaesthesia and hyperbole, with strong, symbolic ironic charge allowing for polysemic interpretations with and without text.

The image action takes place in a car park. Two cars and two Oriental people are visible (Figure 3). One of them is thin, kneeling and trying to break into a BMW and is using a skeleton key to carefully open the door; he has a dark-green balaclava, a case strapped to his belt and he is looking at the car lock. Close to his knees are some tools. The other character, located to the left of would-be assailant, is fat and wearing a red jumpsuit, the only point of figure/ground colour of the image (even though we know that the colour only attracts us in the first two seconds of observation). He is wearing a turban, is stationary, and is taking a side long view at the movements of the would-be assailant and is holding two cymbals in his open arms. The dynamics are minimal but the imminent movement of the two characters is enormous and relevant to understanding the image.



**Figure 3.** Advertising image, from Naga DDB, Malaysia, for car alarms Cobra, 2007. In our study we manipulate the image, cutting off the text or translating it to Portuguese.

### 3.2.3 Participants and apparatus.

Eighteen participants took part in this study, 10 for *Images With Text (IWT)*, 8 for *Images Without Text (IWTT)*. Only 14 registers of eye movement could be considered (7 IWT plus 7 IWTT). All participants observed the image monitored by the ASL 504 model eye tracking system, which has an average output of 60HZ, a precision of 0.5° of visual angle and a visual amplitude of 50° horizontally and 40° vertically. The images were presented as slides using the *Power Point* programme on a 21 inch monitor about 60 cm away from the observer, for a maximum length of 30 seconds. So that the reading of the data supplied by the system could be more rigorous, we divided the image into large areas of interest: the car, the thief, the “alarm” (as the actor in the red jumpsuit has been called), which we divided into two parts (body and cymbals), and the text. The data obtained was treated with the Eyenal programme and exported to Excel to be treated.

### 3.2.4 Procedure.

We registered eye movements during the observation and 48 hours after having shown the advertisement, with and without the text “Cobra Alarms”, we tested the recall and comprehension of the advertisement information.

Three tasks were put to the participants: *a.* to observe the image while being aware that they would be questioned about it; *b.* recalling the contents observed controlled by four questions (1. Identify and characterise the space represented, 2. Identify and characterise the characters in the image, 3. Describe the action (or actions), 4. Identify the elements present in the image apart from the characters) and *c.* information recall, 48 hours later, using the *free recall* model.

### 3.3 Results and discussion for quantity of fixations registered on the text and transition of movements between the text and different areas of interest in the image.

We only discuss here qualitative data from the transitions observed between image and text that we have not subjected to any statistical treatment apart from percentage analysis. The data collected confirm the data observed regarding transitions for captioned images (cf. Baptista, 2009) described above and demand some more experiments with different kind of texts.

#### 3.3.1 Quantity of fixations registered on the text and transition of movements between the text and different areas of interest in the image.

The text was not read first. Even though we know that the text, given its short length, can be read in just one or two fixations, it received on average five fixations (varying between 3 and 9) and the total length of observation was not constant. Fixations on the text are always interspersed with fixations on other areas (Table 3).

**Table 3.** *Duration of fixation sum and average number of fixations*

	Duration of fixation sum		Average number of fixations	
	IWT	IWTT	IWT	IWTT
out	4,879	5,211	15.143	14.857
car	4,595	6,444	10.286	12.857
assailant	4,396	4,239	7.857	8.286
alarm	8,436	7,676	18.714	16.286
text	2,059	—	5.00	—
TOTAL	24,365	23,569	57.000	52.286

This information advocates in favour of a hybrid processing for this type of text, contrary to what has been put forward in previous studies on various types of mixed text, with the exception of captioned images from Baptista (2009) and advertisements from Radach et al (2003), and allows us to presuppose that there are permanent confirmation strategies for previously read material within the area where two textual instances coincide, which could suggest some relation between the sequence of fixations and distrust concerning pictorial text/verbal text coherence and/or doubts in understanding it.

The Average fixation length (afl) on the verbal text registers average values (0.474ms) far above the usual afl values for reading texts, even bearing in mind that this text is written in highly condensed, sans-serif lower case text and with a diminished *track-amount* (which could lead to an increase in afl), even knowing that it could be considered as an iconic-verbal text, given that next to the script “cobra alarms”, there is the head of a cobra. The fact that we have registered transitions between the text and other areas allows us to think that the observer, when confronted with visual information which does not mimetically duplicate the text, might not trust what he is reading and dedicates a significant fixation period to the area of text. In other words, the non-existence in the image of an iconic correspondence for the designation read in the text might lead to confirmation strategies (Table 4).

**Table 4.** *Average number of transitions between interest areas*

	out	car	assailant	alarm	text
out	8,000	1,286	1,000	3,571	1,286
car	1,714	5,571	1,571	0,429	0,571
assailant	0,857	1,000	2,714	2,857	0,429
alarm	2,857	1,714	2,286	10,571	0,714
text	0,714	0,714	0,286	1,286	2,000

Since we know the alarm is represented connotatively (we could say transformed), there is a metonymic dimension between the alarm-noise and the cymbals-noise, between the alarm-alert and the gaze-alert, between the imminent alarm-going off and the imminent cymbals-crashing, but the delirious personification of this alarm displays an attitude of *mimesis phantastikê*, through the cognitive association between the alarm which sets off the musical noise and the music which sets off the alarm noise. This alotopic element, from our point of view, is the richest part of the image. It is this element that makes the rhetorical transposition between the text and image by generating a poetic, transformative, interpretive representation which probably demands transitions between text and image.

### 3.4. Visual processing and understanding hybrid text

The great percentage of fixations, both in IWT (Table 5) and IWTT (Table 6) always occurs in the area considered to be the alarm (with values of 33.806% and 32,251% respectively). The percentage of fixations immediately following is registered precisely outside of any of the areas of interest as defined.

**Table 5.** *Data from Cobra alarms advertisement image observed with text (IWT)*

AOI_Name,	Sum_Fix_Dur,	Fix_Dur_%,	Sum_Fix_Cnt,	Fix_Cnt_%,	Mean_Fix_Dur,
TOTAL	24,365	100,002	57,000	100,001	0,460
OUT	4,879	20,391	15,143	26,394	0,313
CAR	4,595	19,197	10,286	18,191	0,479
ASSAILANT	4,396	17,869	7,857	14,051	0,545
ALARM	8,436	33,806	18,714	32,808	0,488
TEXT	2,059	8,738	5,000	8,559	0,474

In other words, ironically, the area competing with the “alarm” is none of the areas defined as areas of interest. More concretely, a percentage of fixations of 26.394% in IWT and 27.569 in IWTT occurs on the floor, ceiling, columns, walls and the background of the image.

**Table 6.** *Data from Cobra alarms advertisement observed without text (IWTT)*

AOI_Name,	Sum_Fix_Dur,	Fix_Dur_%,	Sum_Fix_Cnt,	Fix_Cnt_%,	Mean_Fix_Dur,
TOTAL	23,569	100,000	52,286	100,000	0,461
OUT	5,211	21,897	14,857	27,569	0,350
CAR	6,444	27,547	12,857	23,904	0,544
ASSAILANT	4,239	18,305	8,286	16,205	0,512
ALARM	7,676	32,251	16,286	32,322	0,437

Data from another research show that the observer’s gaze in most case fixes on alotopic elements, elements outside of the context (which seems right, given that we can ask ourselves why are cymbals being played in a garage?), but it also states that participants do not fix in anyway on the contextualizing elements of the characters in the image, unless given precise instructions to do so. What then leads the participants in this experiment to scan precisely those contextualizing areas? It is not on the red jacket of the musician which participants make the largest number of fixations (not even in the initial two seconds of observation), or on his face, specifically the eyes – as can be predicted when looking at a face – but it is, above all, on the cymbals that the largest number of fixations take place.

Is it that, given the complexity of the image, the theme of the image is not immediately understood in the first 60 ms and a careful examination of the cymbal player, identified as an alotopic element, might correspond to a process of connotative interpretation of the image, making the observer later seek in the shadowed areas, an element which denotatively corresponds to the text?

Does the observer seek, after an initial observation which does not clarify everything, something that can function as a research instruction? Does the text fulfil this function? Does this type of examination of the contextualizing areas of the scene constitute a demand due to the fact that the gaze did not find the mimetic representation for the alarm referred to in the text?

Contrary to what we expected, not all participants in IWTT understood this metonymic figure of the “alarm” character (5 out of 8 understood the situation of the car theft but only 3 out of 8 completely understood the “alarm” character; 3 out of 8 neither understood the “alarm” character nor the car theft and 5 out of 8 did not understand the “alarm”) which seems to advocate in favour of a greater rhetorical complexity than foreseen, but also in favour of the importance of the presence of a designative verbal text, since only 1 participant out of 10, in the IWT, did not understand either the car theft situation or the alarm. This participant presented a very irregular pattern of eye movement and fixations on more distinct points in each area.

Thus we are convinced that this simple designation defines the advertising objectives of those producing the image and clearly demands the understanding of the connotative element as being the only one in the image capable of corresponding to the product being advertised, but also helps to understand the situation of the car theft, which to us seems unequivocal and only was it for those participants with IWT.



Apparently, without a guiding text, focusing on significant elements is always more difficult. Strangely, the participants in the IWTT experiment spent less percentage time on the “alarm” area and spent more time on the shadowed setting than those in IWT.

We conclude that the text attracts attention on the “alarm” character, attributing to this area the possibility of solving the interpretation of the pragmatic dimension of the advert. On the other hand, the absence of text decreases the number of fixations in the shadowed areas, but increases the length of these fixations. In other words, despite making the effort to see, the participant returns to these areas fewer times. It would seem that in the absence of interpreting the “alarm” metonym, the participant turns his attention to other more denotative areas.

## 4. CONCLUSIONS

What attracts in the first place the participants’ attention in this image?

Of the 14 participants 13 first fixed on the cymbals of the *alarm* character. The alotomy referred to was not morphological, based on the chromatic element (fixations on the point of colour of the red shirt) but semantic, on the cymbals (elements outside of the semantic context in a denotative approach). According to previous data – we know that objects with implicit movement attract the attention – this fixation is justified by the imminent movement of the cymbals. It can also be justified by the search of a narrative logic.

But how do the participants structure the transitions between the different elements of the image? Of the 13 participants who first fixed on the alarm, 12 made the transition to the *thief* in second place and just one made the transition to *outside* and then to the *car*. The divergence is minimal and was only present in participants who did not read the text. Even knowing that the human figure always receives the greatest attention, these fixations on the “alarm”, particularly on the cymbals and not his face/eyes, which are expressive and central in the area of observation, and the immediate transition to the theme of *thief* occur with a surprising regularity and indicate that, when faced with an alotopic element, the brain performs in resolving this alotomy, commanding the orientation of eye movement to related elements (syntactically and semantically) with what was first observed.

The greatest number of eye movement transitions from the text is towards the alarm. Nobody looked for the alarm in the car or in the car flashlights; after reading the text, the observer limited himself to confirming the function of transforming representation that the fantastic element of the cymbals has taken on.

The greatest number of transitions from the alarm can be seen to areas outside the areas defined as areas of interest. We can infer that the observer sought a justification for the alotomy which might be hidden and not easily perceptible. The greatest number of eye movement transitions from the *thief* in IWT can be seen towards the *alarm* (an average value of 2.85, followed by an average value of 1.00 for the car, which was in 2<sup>nd</sup> place). In the IWTT participants the greatest number of transitions was also seen towards the *alarm* (an average value of 2.71, but followed by an average value of 2.29 for the car, which shows a much weaker relationship between the *alarm/thief*).

This data advocates, in IWT participants, a search for an interpretation of the connotative relation of the two characters *in the economics of the narrative*. And the logic for the understanding of the narrative in this hybrid advertising text is a connotative sequence that links the thief with the cymbals (the alarm) and not with a common car alarm despite the thief being a car thief. It is this narrative the observer looks for and so the denotative relationship between robber-robbed is not necessary to understand the reason for the presence of the cymbals in the image and it is thus not obvious that the observer needs to look for it. This kind of observation seems to be a result of the hybrid characteristics of the text and so we may probably say that a hybrid text foster hybrid visual processing.

The transitions observed lead us to say that the existence of text, contributing to the designation/interpretation of an alotopic element and not just for its identification, alters the strategies of observation of the image and we can also consider that the presence or absence of verbal text recruits different strategies: in the processing of hybrid texts, the strategy is the confirmation of the meanings between different semiotic systems (setting off the alarm, as a consequence of the thief’s action); in monomodal texts, the strategy is the justification of an interpretation (the “car alarm”).

*Acknowledgements* - This paper is supported by a FCT Project (PTDC/LIN/67854/2006) Comprehension on reading (Bimodal Texts) developed on CLUL and on the FLUL's Psycholinguistics Laboratory.

## 5. REFERENCES

- Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, 103, 62-70.
- Baptista, A., Faria, I., & Luegi, P. (2010). *Hybrid processing of rhetorically elaborated texts*. Proceedings of Communication Cognition and Media in Universidade Católica. Braga (no prelo)
- Baptista, A. (2009). *Interações Texto/Imagem: a caso particular da legenda de fotografia*. Fundação Calouste Gulbenkian /Fundação para a Ciência e Tecnologia.
- Becker, W., & Jurgens, R. (1979). Analysis of the saccadic system by means of double step stimuli. *Vision Research*, 19, 967-983.
- Brandt, H. F. (1945). *The psychology of seeing*. New York: The Philosophical Library.
- Carroll, P. J., Young, J. R., & Guertin, M. S. (1992). Visual analysis of cartoons: a view from the far side. In K. Rayner (Ed.), *Eye movements and visual cognition: scene perception and reading* (pp.444-461). New York: Springer-Verlag.
- Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, 108, 316-355.
- Friedman, A. & Liebelt, L. S. (1981). On the time course of viewing pictures with a view towards remembering. In D. F. Fisher, R. A. Monty & J. W. Senders (Eds.), *Eye Movements: Cognition and Visual Perception* (pp.145-168). Hillsdale, NJ: Erlbaum.
- Hegarty, M. (1992a). The mechanisms of comprehension and comprehension of mechanisms. In K. Rayner (Ed.), *Eye movements and visual cognition: scene perception and reading* (pp.428-443). New York: Springer-Verlag.
- Hegarty, M. (1992b). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 1084-1102.
- Faria, I. H., Luegi, P., Taborda, C., & Baptista, A. (2006). Recuperação da Informação visualizada: Interação e Competição entre legendas e imagens. In *Actas do XXI Encontro da Associação Portuguesa de Linguística* 2005. Porto: FLUP
- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive Determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 565-572.
- Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics*, 2, 547-552.
- Mitchell, W. J. T. (1994). *Picture theory: Essays on verbal and visual representation*. Chicago: The University of Chicago Press.
- Radach, R., Lemmer, S., Vorstius, C., Heller, D., & Radach, K. (2003). Eye movements in the processing of print advertisements. *The Minds Eye: Cognitive and Applied Aspects of Eye Movement Research*. Amsterdam: Elsevier Science BV.
- Rayner, K., & Pollatsek, A. (1989). *The Psychology of reading*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Rayner, K., Rotello, C., Stewart, J. A., Keir, J. & Duffy, A. S. (2001). Integrating Text and Pictorial Information: Eye Movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, 7(3), 219-226.

- Rayner, K., & Castelhana, M. S. (2007). Eye movements during reading, scene perception, visual search and while looking at print advertisements. In M. Wedel & R. Pieters (Eds), *Visual marketing: from attention to action* (pp. 9-42). New Jersey: Lawrence Erlbaum.
- Rayner K. Miller B., & Rotello, C. (2008). Eye Movements When Looking at Print Advertisements: The Goal of the Viewer Matters. *Applied Cognitive Psychology*, 22(5), 697-707.
- Scherner, M. (2000). Texto: Subsídios para a história do conceito. In I. B. Duarte, F. Henriques & I. M. Dias (Orgs.), *Texto leitura e escrita: Antologia* (pp.125-181). Porto: Porto Editora. (Trad. port. de Helder Lourenço de Text. Untersuchungen zur Begriffsgeschichte, 1996).
- Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, 59 (11), 1931 – 1949.